

<https://helda.helsinki.fi>

---

## Interaction of the effects associated with auditory-motor integration and attention-engaging listening tasks

Wikman, Patrik

2019-02

---

Wikman , P & Rinne , T 2019 , ' Interaction of the effects associated with auditory-motor integration and attention-engaging listening tasks ' , Neuropsychologia , vol. 124 , pp. 322-336 . <https://doi.org/10.1016/j.neuropsychologia.2018.11.006>

---

<http://hdl.handle.net/10138/318371>

<https://doi.org/10.1016/j.neuropsychologia.2018.11.006>

---

cc\_by\_nc\_nd

acceptedVersion

---

*Downloaded from Helda, University of Helsinki institutional repository.*

*This is an electronic reprint of the original article.*

*This reprint may differ from the original in pagination and typographic detail.*

*Please cite the original version.*

1 **Interaction of the effects associated with auditory-motor integration**  
2 **and attention-engaging listening tasks**

3  
4 Patrik Wikman<sup>1,2</sup> and Teemu Rinne<sup>1,3</sup>

5  
6 <sup>1</sup>Department of Psychology and Logopedics, University of Helsinki, Helsinki, Finland

7 <sup>2</sup>Advanced Magnetic Imaging Centre, Aalto University School of Science, Espoo,  
8 Finland

9 <sup>3</sup>Turku Brain and Mind Center, Department of Clinical Medicine, University of  
10 Turku, Turku, Finland

11  
12  
13 **Corresponding author:** Patrik Wikman, Department of Psychology and Logopedics,  
14 PO Box 9, FI-00014 University of Helsinki, Phone: +358 294129533, email:  
15 patrik.wikman@helsinki.fi

16 **Keywords:** speech repetition, speech production, attention, auditory cortex, inferior  
17 parietal lobule

## Abstract

A number of previous studies have implicated regions in posterior auditory cortex (AC) in auditory-motor integration during speech production. Other studies, in turn, have shown that activation in AC and adjacent regions in the inferior parietal lobule (IPL) is strongly modulated during active listening and depends on task requirements. The present fMRI study investigated whether auditory-motor effects interact with those related to active listening tasks in AC and IPL. In separate task blocks, our subjects performed either auditory discrimination or 2-back memory tasks on phonemic or nonphonemic vowels. They responded to targets by either overtly repeating the last vowel of a target pair, overtly producing a given response vowel, or by pressing a response button. We hypothesized that the requirements for auditory-motor integration, and the associated activation, would be stronger during repetition than production responses and during repetition of nonphonemic than phonemic vowels. We also hypothesized that if auditory-motor effects are independent of task-dependent modulations, then the auditory-motor effects should not differ during discrimination and 2-back tasks. We found that activation in AC and IPL was significantly modulated by task (discrimination vs. 2-back), vocal-response type (repetition vs. production), and motor-response type (vocal vs. button). Motor-response and task effects interacted in IPL but not in AC. Overall, the results support the view that regions in posterior AC are important in auditory-motor integration. However, the present study shows that activation in wide AC and IPL regions is modulated by the motor requirements of active listening tasks in a more general manner. Further, the results suggest that activation modulations in AC associated with

- 42 attention-engaging listening tasks and those associated with auditory-motor
- 43 performance are mediated by independent mechanisms.

## 1 Introduction

Auditory sensory feedback is used to monitor the accuracy of produced speech. For example, many studies have shown that real-time pitch shifting of one's own voice results in articulatory changes in the opposite direction to compensate for the artificial shift (Burnett et al., 1998; Purcell and Munhall, 2006; Tourville et al., 2008). Previous research using fMRI has implicated the regions at the boundary of left temporal and parietal cortices in the posterior planum temporale (pPT) as important for interfacing between auditory and motor information (Hickok et al., 2011; Hickok et al., 2009; Okada et al., 2010; Peschke et al., 2012; Peschke et al., 2009; Shuster and Lemieux, 2005; Simmonds et al., 2014b). This idea is supported by results showing enhanced activation in this region during both listening to speech and covert speech production as well as during non-speech auditory-motor tasks such as humming of melodies or playing a musical instrument (Buchsbaum et al., 2001; Hickok et al., 2003a; Hickok et al., 2009; Pa and Hickok, 2008). Left pPT has been suggested to be particularly important for actions that are novel and nonautomatic such as repetition of vocalizations made by other individuals (Hickok, 2012; Hickok, 2016). Accordingly, damage in this region has been associated with a selective difficulty in repeating heard words (i.e., conduction aphasia, Baldo et al., 2008; Buchsbaum et al., 2011; Rogalsky et al., 2015) with otherwise intact speech perception or production. Regions in bilateral inferior parietal lobule (IPL), in turn, have been implicated in processing of somatosensory feedback during speech production (Golfinopoulos et al., 2011; Guenther, 2006; Guenther et al., 2006; Guenther and Vladusich, 2012; Hickok, 2012; Tremblay et al., 2003).

Posterior superior temporal gyrus (STG) and IPL regions also show strong activation modulations that depend on the characteristics of the listening task. Attention-engaging auditory tasks, in general, are associated with enhanced activation in wide regions of auditory cortex (AC) along STG (Alho et al., 2014; Hall et al., 2000; Petkov et al., 2004; Rinne, 2010; Rinne et al., 2017; Rinne et al., 2005; Woods and Alain, 2009). Further, regions in anterior-middle STG show higher activation during (acoustical) discrimination than during (categorical) n-back memory tasks performed on identical stimuli, whereas during n-back tasks activation is higher in IPL (Harinen and Rinne, 2013; Harinen and Rinne, 2014; Häkkinen et al., 2015; Rinne et al., 2009; Rinne et al., 2012). Such attention- and task-related modulations could easily interact with the effects associated with sensory-motor integration seen in homologous regions in posterior STG and IPL. However, to our knowledge, previous studies have not systematically investigated the links between the modulations during attention-engaging listening tasks and those associated with auditory-motor integration.

In the present fMRI study (blocked design), we tested whether task and auditory-motor effects interact in AC and adjacent regions. To this end, subjects performed demanding vowel discrimination and vowel-category 2-back memory tasks and responded to targets using overt vocalizations or button presses (Figure 1 and Table 1). In the discrimination task, subjects were required to indicate when the two vowels of a pair were acoustically identical. During the 2-back task, in turn, subjects indicated when a vowel pair belonged to the same vowel category as the one presented two trials before. The vocalization responses were made either by repeating the last vowel of a target pair (vocalization based on direct auditory-to-motor translation) or by producing a given response vowel (vocalization not based on

auditory input). In different blocks, the vowels were either phonemic (Finnish) or nonphonemic (i.e., not a Finnish phoneme) vowels. In addition to these vowel tasks, there were also analogous pitch discrimination and pitch-category 2-back tasks performed on pitch-varying vowels.

This experimental design allowed us to investigate the effect of task (discrimination vs. 2-back tasks), motor-response type (vocal vs. button responses), vocal-response type (vocal repetition vs. production), and vowel-type (phonemic vs. nonphonemic vowels) effects as well as their interactions. Motor-response type was varied in order to investigate the unknown activation differences associated with vocal and button responding. Our previous study showed that activation in AC and IPL strongly depends on whether subjects respond to targets in a listening task using manual responses or without manual responses (Wikman et al., 2015). Effects associated with motor responding, if not controlled for, could easily affect the interpretation of results obtained in auditory-motor tasks. Vocal-response type, in turn, was varied to measure auditory-motor integration effects. Previous literature suggests that repetition of utterances made by other individuals requires translation of acoustic input to motor programs in posterior STG (i.e. auditory-motor integration; Baldo et al., 2008; Buchsbaum et al., 2011; Rogalsky et al., 2015; Parker Jones et al., 2014; Simmonds et al., 2014), whereas a self-selected vowel can be produced based on a more direct activation of the motor programs (i.e. less load on auditory-motor integration; Hickok, 2012; Hickok, 2016). Therefore, auditory-motor integration effects should be stronger during repetition than production responses. In the present study, auditory-motor integration was tested also by varying the vowel type. The requirements for auditory-integration should be higher during repetition of nonphonemic than phonemic vowels, as repetition of nonphonemic vowels relies

118 more on a direct translation of acoustic input to motor commands, whereas repetition  
119 of phonemic vowels benefits from the well-learned representations of native language  
120 (Wilson and Iacoboni, 2006). We also aimed to investigate whether auditory-motor  
121 integration effects are similarly observed in tasks requiring vowel repetition or pitch  
122 repetition (humming).

123               Specifically, in the present study, we tested the following main  
124 hypotheses regarding auditory-motor integration: (H1) Due to greater requirements  
125 for auditory-motor integration, vowel repetition responses should be associated with  
126 stronger activation in AC, particularly in pPT, than vowel production responses. (H2)  
127 Similarly, requirements for auditory-motor integration, and the associated  
128 modulations, should be higher during repetition of nonphonemic than phonemic  
129 vowels. (H3) Further, if auditory-motor and task-dependent effects are independent of  
130 each other, then similar auditory-motor effects should be observed during both  
131 discrimination and 2-back tasks. However, if this is not the case, then auditory-motor  
132 effects could be, at least partially, related to changes in task demands rather than to  
133 auditory-motor processing as such.



## 2 Materials and Methods

### 2.1 Subjects

Twenty healthy subjects (12 women; age 18–28, mean 24; no known hearing deficits; all right handed) participated after providing informed consent. The study protocol was approved by the University of Helsinki Ethical Review Board in the Humanities and Social and Behavioural Sciences.

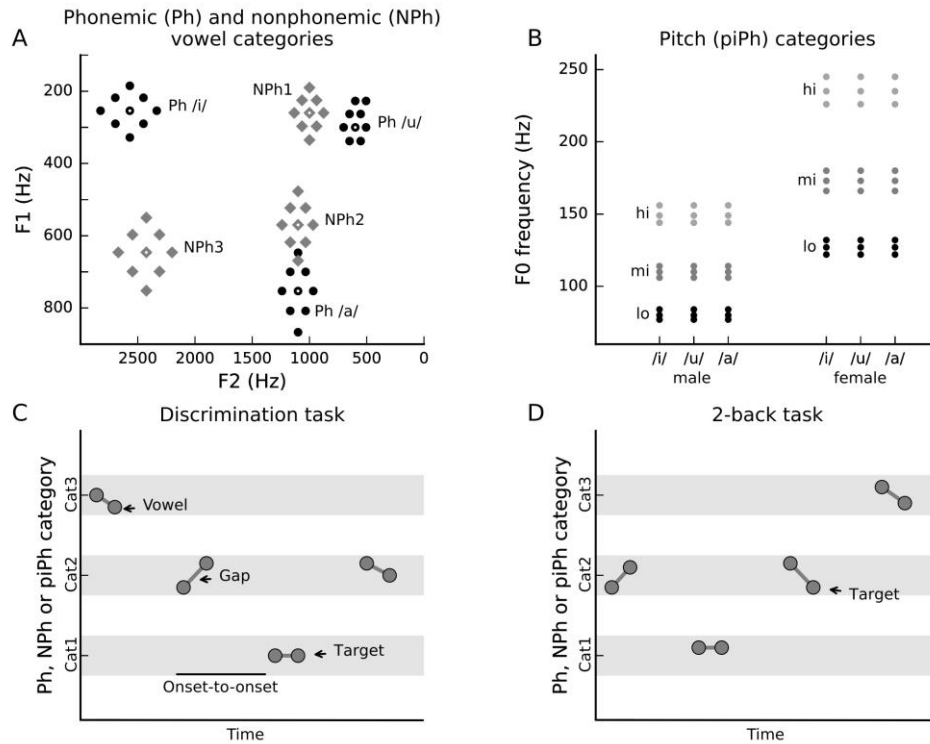
### 2.2 Stimuli

The stimuli were Finnish phonemic (Ph) and nonphonemic (NPh) vowels synthesized using the Praat software package (version 5.1.12, [www.praat.org](http://www.praat.org)). The Ph and NPh vowels and categories were identical to the ones used in our previous study (Harinen and Rinne, 2013). In brief, there were three Ph and three NPh vowel categories with nine vowels in each (Fig. 1 A). The three Ph categories were defined in F1-F2 space around typical Finnish /i/, /u/ or /a/ vowels. The NPh categories (NPh1, NPh2 and NPh3) were defined in regions of F1-F2 space where no Finnish phonemes exist. Within each category, the vowels were separated by at least 60 mel. The F1 and F2 frequencies for the center phoneme (indicated with a white dot in Fig. 1 A) of each category were as follows: /i/ 254 and 2569 Hz, /u/ 300 and 600 Hz, /a/ 753 and 1100 Hz, NPh1 260 and 1000 Hz, NPh2 570 and 1100 Hz, and NPh3 646 and 2425 Hz. A linear falling contour from 150 to 100 Hz was used for the F0. All vowels were 200 ms in duration (including a linear 5 ms onset and offset ramp).

In addition to fixed-pitch Ph and NPh vowels, there were low, medium and high pitch Ph (piPh) vowel categories (separated by 4 semitones) each containing

three different vowel sounds with three pitch levels (separated by 0.7 semitones, Fig. 1 B). These vowels were otherwise identical to the center phonemes of each Ph category, but a constant F0 was used to create a distinct pitch. Slightly lower F0 values were used for male (low: 77, 80 and 84 Hz; medium: 106, 110 and 114; high: 144, 149 and 156 Hz) than female (low: 122, 127 and 132 Hz; medium: 166, 173 and 180 Hz; high: 226, 235 and 245 Hz) subjects.

The vowels (200 ms in duration) were presented in pairs where the first and second vowel were separated by a 100 ms gap. The vowel pairs were presented with 1400–1900 ms onset-to-onset intervals (i.e., from the onset of a pair to the onset of the next pair). In each pair, both vowels belonged to the same Ph, NPh or pitch vowel category (Fig. 1 C–D). The Ph and NPh vowel pairs consisted of either two identical vowels or two different vowels within the same vowel category separated by 60 mel. The piPh vowel pairs consisted of two identical vowels within the same pitch category that were either identical in pitch or separated by 0.7 semitones. In one task block, all stimuli were either Ph, NPh or piPh vowel pairs.



**Fig. 1. Stimuli and tasks.** (A) In vowel task blocks, subjects were presented with vowel pairs from three phonemic (Ph, black circles) or three nonphonemic (NPh, gray diamonds) vowel categories, each with nine different vowels. The Ph categories were defined based on typical Finnish /i/, /u/ and /a/ phonemes (white dot). The NPh categories were organized in regions of F1-F2 space where no prototypical vowels exist. (B) Three pitch categories (low, middle and high) were defined separately for male and female subjects. Each pitch category contained nine different sounds (three different vowels, three pitch levels). The pitch-modulated /i/, /u/ and /a/ vowels (piPh) were created based on the center-of-category Ph vowels indicated by a white dot in panel A. (C, D) The vowels were presented as within-category Ph (e.g., /i<sub>2</sub>–/i<sub>7</sub>), NPh (e.g., NPh<sub>13</sub>–NPh<sub>1</sub>) or piPh (e.g., male low /i<sub>1</sub>–/i<sub>3</sub>) pairs. Note that the vowels in piPh pairs could differ only in pitch. (C) In the discrimination task, subjects were required to indicate when the first and the second part of the pair were acoustically identical. (D) In the vowel 2-back task performed on Ph or NPh vowel pairs, subjects

indicated when the pair belonged to the same vowel category as the one presented two trials before. With piPh vowel pairs, subjects performed the same task based on pitch categories (vowel identity was task irrelevant).

### *2.3 Tasks and responses*

The vowel pairs were presented during discrimination or 2-back task blocks. In the vowel discrimination tasks (Fig. 1 C), subjects were required to respond when the first and second part of the vowel pair were identical. In the 2-back task (Fig. 1 D), subjects responded when the vowel pair belonged to the same (Ph, NPh or piPh) category as the sound pair presented two trials before. Note that the 2-back task with piPh vowel pairs was performed based on pitch categories (low, middle, high) and that in this task vowel identity was task irrelevant. In addition, a simple count task was presented as an additional (easy) control condition. In the count task, subjects were required to respond to every fourth pair.

In different blocks, subjects responded either by pressing a response button with their right index finger (button response blocks; Bu), by overtly vocalizing a given response phoneme (/æ/, /o/ or /y/; phoneme production response blocks, phPr), or by overtly vocalizing the last part of the target pair (phoneme or nophoneme repetition response blocks, phRe or nphRe). In repetition-response blocks with pitch-modulated vowels, subjects hummed the pitch of the second vowel of the target pair (pitch repetition response blocks, piRe). In the count task, only repetition and button responses were used.

The combination of task (discrimination, 2-back, count), stimulus type (Ph, NPh, piPh) and response type (Bu, Pr, Re) resulted in 18 different conditions.

Together with the count task (Ph, NPh or piPh stimuli; repetition or button responses), there were 24 conditions altogether (Table 1).

The experimental conditions were presented in 12.5 s blocks with 7 vowel pairs in each alternating with 12.5 s breaks with no stimuli. During the breaks, subjects focused on a fixation mark presented in the middle of a visual display. A graphic task instruction symbol replaced the fixation mark 2.5 s before the beginning of the next task block. The instruction symbol indicated the task, the type of stimuli and the type of response. In phPr blocks, the response vowel was indicated in the middle of the instruction symbol. The graphic instruction symbol was presented until the end of the task block. Each task block contained 2–3 target pairs (15 target pairs per condition). Each subject completed two runs with 72 task blocks in each (i.e., 6 blocks per condition; the order of task blocks was randomized within a run; half of subjects started with run 1 and half with run 2).

The auditory stimuli were delivered using Sensimetrics S14 insert earphones (<http://www.sens.com/>, Malden, USA). The noise of the scanner was attenuated by the insert earphones, circumaural ear protectors (Bilsom Mach 1) and viscous foam pads attached to the sides of the headcoil. The graphic instruction symbols were presented in the middle of the screen via a mirror fixed to the head coil. The experiment was controlled using the Presentation software (Neurobehavioral Systems, Berkeley, CA, USA)

**Table 1.** Summary of experimental conditions (blocked design). Phonemic (Ph), nonphonemic (NPh) or pitch-varying phonemic (piPh) vowel pairs were presented during discrimination, 2-back or count task blocks. In these tasks, subjects responded either by repeating the last part of a target pair (phRe, nphRe), humming its pitch (piRe), producing a given response vowel (phPr), or pressing a button (Bu). There were nine different discrimination, nine 2-back, and six count conditions (i.e., 24 conditions in total).

Task	Vowel	Response
Discrimination or 2-back	Ph	phRe, phPr, Bu
	NPh	nphRe, phPr, Bu
	piPh	piRe, phPr, Bu
Count	Ph	phRe, Bu
	NPh	nphRe, Bu
	piPh	piRe, Bu

#### 2.4 Pre-fMRI training

Before fMRI, each subject was carefully trained to perform the tasks in two training sessions (3 h in total) until they and study personnel felt confident in subjects' ability to properly identify the graphic task-instruction symbols, to correctly perform the corresponding (and demanding) tasks, and perform the button and vocal responses without excessive head movements.

#### 2.5 Analysis of behavioral performance

During fMRI, the vocal responses were recorded using an MRI compatible microphone (FOMRI, Optoacustics, Or Yehuda, Israel). The onsets of the vocalizations were identified using in-house Python scripts and manually verified. Vocal and button responses occurring between 200 and 1300 ms from the onset of the

target stimulus were accepted as hits. Other responses (i.e., responses not within the response window) were considered as false alarms. Hit rate (HR) was defined as the number of hits divided by the number of targets. False alarm rate (FaR) was defined as the number of false alarms divided by the number of nontargets. HRs and FaRs were calculated separately for each experimental condition. Mean HR and FaR were used to compute  $d'$  (index of stimulus detectability,  $d' = [Z(\text{HR}) - Z(\text{FaR})]$ ) and response bias ( $c = -0.5 * [Z(\text{HR}) + Z(\text{FaR})]$ ). Reaction times were only calculated for hits.

## *2.6 fMRI data acquisition and analysis*

Functional magnetic resonance imaging data were acquired with a 3T MAGNETOM Skyra scanner (Siemens Healthcare, Erlangen, Germany) using a standard 20-channel head-neck coil. First, a high-resolution anatomical image (sagittal slices, slice thickness 1.0 mm, in-plane resolution  $1.0 \times 1.0 \text{ mm}^2$ ) was acquired. Second, functional images (GE-EPI; TR 2200 ms, TE 30 ms, flip angle  $78^\circ$ , voxel matrix  $96 \times 96$ , FOV 18.9 cm, slice thickness 2.0 mm with no gap, in-plane resolution  $2.0 \times 2.0 \text{ mm}^2$ , 29 slices) were acquired in two 34 min runs. The middle EPI slices were aligned along the Sylvian fissures based on the anatomical image. The imaged area covered the superior temporal lobe, insula and most of the inferior parietal lobes in both hemispheres (Fig. 2A). Finally, a T2-weighted image using the same imaging slices but a denser in-plane resolution was acquired (TR 4500 ms, TE 100 ms, voxel matrix  $256 \times 256$ , FOV 18.9 cm, slice thickness 2.0 mm).

Cortical surface reconstruction and coregistration were conducted using Freesurfer (version 5.3, [www.freesurfer.net](http://www.freesurfer.net)). Functional data were motion corrected, resampled to the standard cortical surface, and spatially smoothed (i.e., in surface

space; 10 mm FWHM). Global voxel-wise analysis was performed in surface space using FSL's (version 6.0; [www.fmrib.ox.ac.uk/fsl](http://www.fmrib.ox.ac.uk/fsl)) general linear model. Separate explanatory variables were used for each of the 24 conditions. The hemodynamic response function was modeled with a gamma function (mean lag 6 s, SD 3 s) and its temporal derivative. A second-level analysis using fixed effects was used to combine the results of the two runs. Third level group analysis was performed using PALM (Permutation Analysis of Linear Models, version alpha26, Winkler, Ridgway, Webster, Smith, & Nichols, 2014). Significance was assessed by permutation inference (10 000 permutations; each subject defined an exchangeability block). Correction for multiple comparisons (family-wise error rate, FWER) was performed using cluster mass correction (PALM; initial cluster forming threshold  $Z > 2.3$ ). For visualization, results were converted to 2D using Mollweide projection.

## *2.7 Regions of interest (ROIs)*

Before data analysis, four anatomical regions of interest (Heschl's gyrus HG, anterior planum temporale aPT, posterior planum temporale pPT, and supramarginal gyrus SMG; Fig. 2 E) were defined on the standard cortical surface (fsaverage). The ROIs were hand-drawn based on anatomical landmarks. The HG ROI was defined as in our previous study (Wikman et al., 2015). PT was divided in two ROIs (aPT and pPT) as particularly pPT is implicated in auditory-motor integration (Hickock, Saberi 2012). The SMG ROI was defined based on the sulci separating SMG from the superior parietal lobule and the angular gyrus.



## 2.8 Analysis of the Ph and NPh vowels uttered during fMRI

To verify that subjects were able to repeat the Ph and NPh vowels as instructed, the utterances recorded during fMRI were presented to nine naïve native-Finnish listeners. The utterances of six fMRI subjects were excluded, as the overall quality of the recording was poor. The utterances were played back to the listeners in pairs in which one utterance was from phRe blocks and the other from nphRe blocks. Each utterance was paired with five different utterances of the opposite category. The listeners were required to indicate by pressing one of two buttons whether the first or second vowel of the pair was a Finnish phoneme. The utterances of each fMRI subjects were rated by at least three listeners. That is, each listener classified the utterances of 3–4 fMRI subjects (session duration 30 min). After this procedure, each utterance was associated with a mean classification accuracy (0–100 %) across the three listeners.

## 2.9 Additional fMRI analyses

In addition to the full-block analysis (described in section 2.6.), two additional analyses were conducted to compare (1) effects during the first seconds of each block (2–12 seconds depending on the block) before the first response (block-start analysis) and (2) effects associated with the first response of each block (first-response analysis). The block-start analysis was conducted to test whether activation during vocal-response blocks (Fig. 2 B and 4, Table 4) was influenced by stimulus-dependent effects to self-produced vocalizations (i.e. subjects heard their own voice). The first-response analysis, in turn, tested whether activation during production-response blocks was influenced by the fact that the same response vowel was repeatedly uttered, whereas in repetition-response blocks the response vowel varied

328 (i.e. the magnitude of fMRI signal might decrease with repetition; Bergerbest et al.,  
329 2004).

### 3 Results

#### 3.1 Task performance

Subjects successfully performed the demanding tasks during fMRI. Mean  $d'$ ,  $c$  and RT in each task are shown in Table 2. Note that RTs were always longer for vocal than button responses probably because vocal responding is slower than button responding (Prosek et al., 1979) and RTs measured based on the recorded vocalizations are likely to be associated with an additional delay. Thus, RTs in vocal- and button-response blocks should not be compared with each other. Performance in the vowel, piPh and count tasks was analyzed separately using three- and two-way repeated measures ANOVAs (Table 3).

**Table 2.** Mean  $d'$ ,  $c$ , and reaction time (RT) in each condition. SEM is given in parentheses.

	$d'$			$c$			RT (s)		
	Re	Pr	Bu	Re	Pr	Bu	Re	Pr	Bu
Discrimination									
Ph	1.1 (0.1)	1.5 (0.2)	1.7 (0.2)	0.6 (0.1)	0.7 (0.1)	0.3 (0.1)	1.3 (0.1)	1.3 (0.1)	1.0 (0.1)
NPh	1.6 (0.2)	1.7 (0.1)	1.8 (0.2)	0.7 (0.1)	0.5 (0.1)	0.4 (0.1)	1.3 (0.1)	1.2 (0.1)	1.0 (0.1)
piPh	1.4 (0.3)	1.7 (0.3)	2.2 (0.3)	0.7 (0.1)	0.6 (0.1)	0.4 (0.1)	1.2 (0.1)	1.2 (0.1)	1.0 (0.1)
2-back									
Ph	2.1 (0.2)	2.2 (0.2)	2.6 (0.2)	0.7 (0.1)	0.7 (0.1)	0.5 (0.1)	1.2 (0.1)	1.2 (0.1)	0.9 (0.1)
NPh	2.1 (0.2)	2.0 (0.2)	2.9 (0.3)	0.6 (0.1)	0.8 (0.1)	0.3 (0.1)	1.2 (0.1)	1.2 (0.1)	0.9 (0.1)
piPh	1.2 (0.3)	1.0 (0.2)	1.5 (0.2)	0.8 (0.1)	0.8 (0.1)	0.5 (0.1)	1.2 (0.1)	1.1 (0.1)	1.0 (0.1)
Count									
Ph	4.3 (0.1)	-	4.2 (0.1)	0.1 (0.1)	-	0.1 (0.1)	1.2 (0.1)	-	0.8 (0.1)
NPh	4.3 (0.1)	-	4.1 (0.2)	0.1 (0.1)	-	0.1 (0.1)	1.2 (0.1)	-	0.8 (0.1)
piPh	3.9 (0.2)	-	4.3 (0.1)	0.1 (0.1)	-	0.1 (0.1)	1.1 (0.1)	-	0.8 (0.1)

**Table 3.** The results of three-way and two-way repeated measures ANOVAs on the performance data (N = 20, FDR corrected within and across ANOVAs). Rows with significant (P < 0.05, bold) effects are listed first.

	d'			c		RT	
	df	F	P	F	P	F	P
<b>Vowel tasks: task (Discr. 2-back) × vowel type (Ph. NPh) × motor-response type (Re. Bu)</b>							
task	1,19	12	<b>0.021</b>	0	0.927	32	<b>0.001</b>
motor-response type	1,19	12	<b>0.015</b>	34	<b>0.001</b>	104	<b>0.001</b>
task × vowel type	1,19	3.2	0.276	13	<b>0.015</b>	0.3	0.760
vowel type	1,19	1.3	0.452	1.2	0.459	3.8	0.243
vowel type × motor-response type	1,19	1.2	0.459	1.4	0.414	2.7	0.320
task × motor-response type	1,19	0	0.972	0	0.969	3.5	0.245
task × vowel type × motor-response type	1,19	4.4	0.191	2.0	0.351	0.2	0.784
<b>Vowel tasks: task (Discr. 2-back) × vowel type (Ph. NPh) × vocal-response type (Re. Pr)</b>							
task	1,19	7.8	0.058	1.5	0.405	42	<b>0.001</b>
vowel type	1,19	1.8	0.377	0.2	0.784	0	0.988
vocal-response type	1,19	2.1	0.351	0	0.988	0.5	0.643
task × vowel type	1,19	5.6	0.120	0.2	0.784	1.1	0.459
vowel type × vocal-response type	1,19	2.1	0.351	1.6	0.405	0.2	0.784
task × vocal-response type	1,19	6.4	0.092	8.6	0.052	0.9	0.493
task × vowel type × vocal-response type	1,19	0.7	0.554	6.5	0.092	2.6	0.326
<b>Pitch tasks: task (Discr. 2-back) × motor-response type (Re. Bu)</b>							
motor-response type	1,19	9.0	<b>0.044</b>	13	<b>0.015</b>	19	<b>0.001</b>
task	1,19	1.9	0.359	0.5	0.648	1.1	0.459
task × motor-response type	1,19	2.4	0.328	0	0.970	0.1	0.887
<b>Pitch tasks: task (Discr. 2-back) × vocal-response type (Re. Pr)</b>							
vocal-response type	1,19	0.1	0.833	1.0	0.490	0.4	0.716
task	1,19	2.4	0.328	3.6	0.288	3.6	0.244
task × vocal-response type	1,19	5.4	0.126	3.0	0.244	2.1	0.351
<b>Count task: vowel type (Ph. NPh. piPh) × motor-response type (Re. Bu)</b>							
motor-response type	2,38	0.1	0.784	2.5	0.328	67	<b>0.001</b>
vowel type	1,19	2.0	0.328	1.5	0.405	0.1	0.969
vowel type × motor-response type	2,38	1.5	0.405	1.1	0.501	0.7	0.673

### 3.2 fMRI

First, we investigated the overall effects in AC and adjacent regions using two separate two-way ANOVAs with factors task (discrimination, 2-back) and motor-response type [(repetition, button) or (production, button); (Re, Bu) or (Pr, Bu)]. Both ANOVAs showed that activation in wide regions extending from the insula to STG

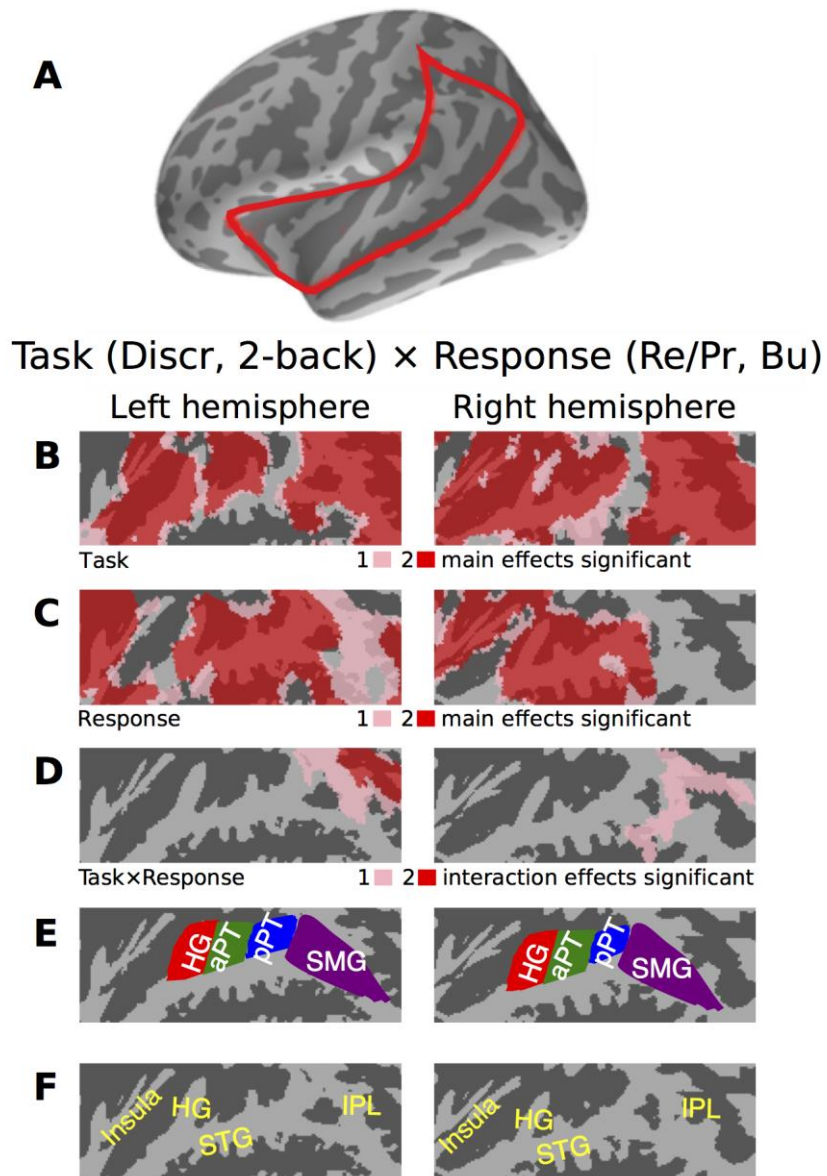
and IPL depended on the listening task (main effects of task, Fig. 2 B). These task-dependent modulations were caused by higher activation in the insula and STG during discrimination than 2-back tasks, whereas the 2-back tasks were associated with stronger activation in IPL. Moreover, both ANOVAs also showed significant main effects of motor-response type in wide STG and IPL regions (Fig. 2 C). These motor-response type main effects were caused by stronger activation during vocal- (Re or Pr) than button-response blocks. In both ANOVAs, task  $\times$  motor-response type interactions were observed in IPL (Fig. 2 D), where activation was stronger during vocal-response blocks than during button-response blocks in the discrimination but not in the 2-back tasks.

The overall effect of vocal-response type was tested with an analogous ANOVA with factors task (discrimination, 2-back) and vocal-response type (Re, Pr). The main effect of vocal-response type was significant in the left pPT and left IPL (see next paragraph) but no significant task  $\times$  vocal-response type interactions were observed.

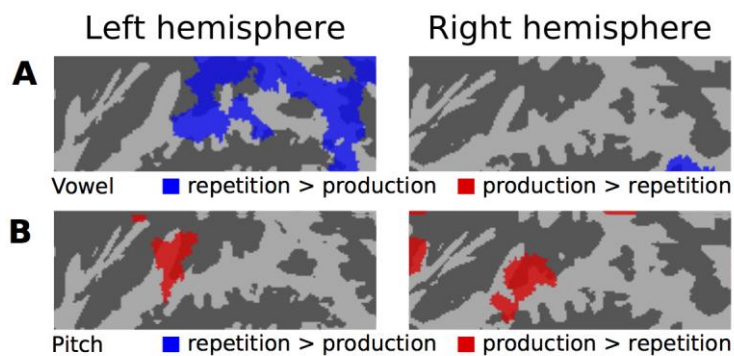
Direct comparisons of activation during the vowel task (Ph and NPh) blocks with repetition and production responses is shown in Fig. 3 A. Activation in left PT and IPL was significantly stronger during vowel repetition than during production response blocks. However, the pitch task blocks with repetition responses, in which subjects responded to targets by repeating the pitch of vowels by humming, were not associated with stronger activation than pitch-task blocks with vowel-production responses. By contrast, activation in the pitch-task blocks was stronger with vowel-production than pitch-repetition responses bilaterally in and near HG (Fig. 3 B).

Comparisons (not shown) between task blocks with Ph or NPh vowels

382 revealed significantly higher activation in STG and IPL regions during NPh blocks.  
 383 This effect was systematically observed only during button-response blocks (cf. Figs.  
 384 4 and 5).



**Fig. 2.** The effects of task and motor-response type on activation in AC and adjacent regions (Ph, NPh and piPh sound blocks collapsed together). (A) The red outline illustrates the imaged area on an inflated left-hemisphere cortical surface (light gray, gyri; dark gray, sulci). (B–D) Results of two task  $\times$  response ANOVAs, in which the task factor (Discr, 2-back) was the same but the response factor varied [(repetition, button) or (production, button); (Re, Bu) or (Pr, Bu)]. The results of the two separate ANOVAs are plotted so that areas where one main effect (B, C) or interaction (D) was significant ( $N = 20$ , permutation inference, FWER corrected  $P < 0.05$ ) are shown in pink and areas where both effects were significant are plotted in red. The results are plotted on flattened two-dimensional maps (average of the 20 subjects) of the imaged area shown in A. (E) Areas used in the ROI analysis: HG Heschl's gyrus, aPT anterior planum temporale, pPT posterior planum temporale, SMG supramarginal gyrus. (F) Anatomical landmarks. STG superior temporal gyrus, HG Heschl's gyrus, IPL inferior parietal lobule.



**Fig. 3.** Direct contrasts between repetition- and production-response blocks ( $N = 20$ , permutation inference, FWER corrected  $P < 0.05$ ). (A) During vowel tasks (collapsed across Ph and NPh blocks), activation in regions of left PT and IPL was stronger when the tasks were performed with repetition than production responses (blue). (B) During pitch tasks, activation in HG and anterior PT regions was stronger during

vowel-production (Ph) than when subjects responded by humming the pitch of the target (red).

### *3.3 ROI analysis*

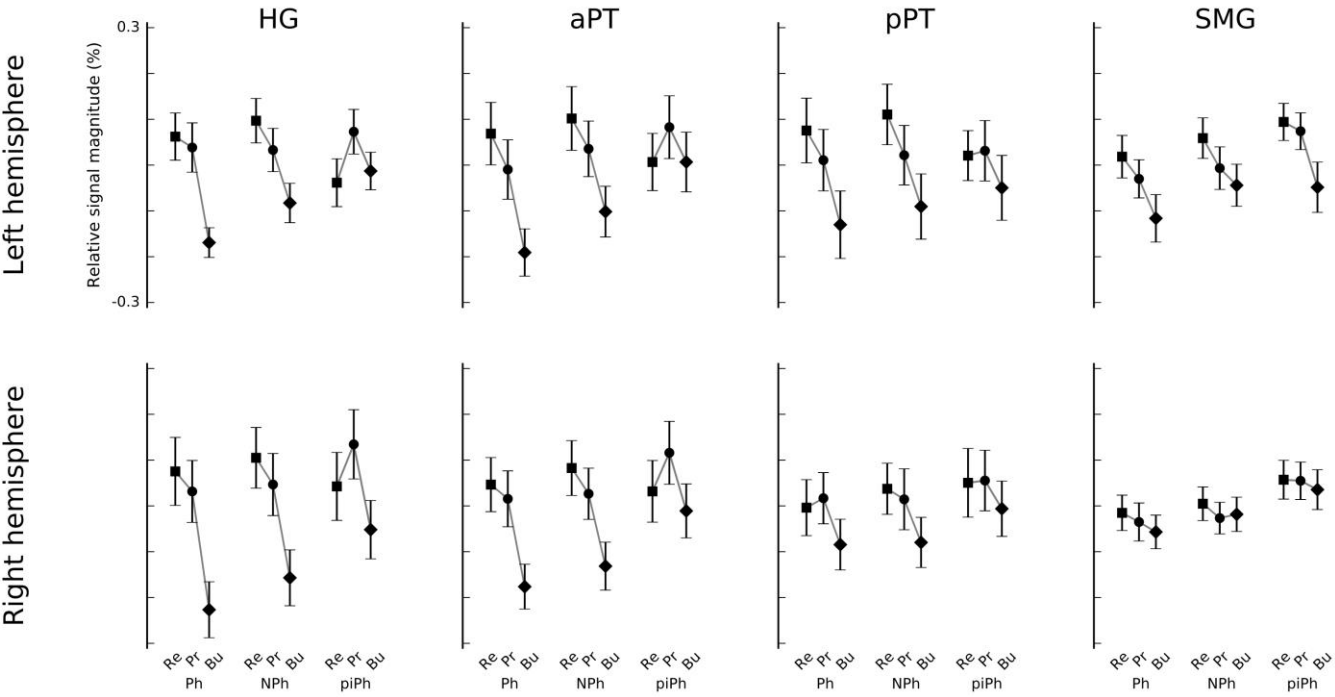
To investigate the motor-response type effects in more detail, we extracted mean signal magnitudes for each condition in four anatomically defined ROIs (Fig. 2 E). We first conducted an omnibus ANOVA with factors hemisphere, ROI (HG, aPT, pPT, SMG), task (Discrimination, 2-back), stimulus type (Ph, NPh, piPh), and response type (Re, Pr, Bu). FDR corrected (within and across ANOVAs) effects are reported in Table 4. Notably, consistent with results in Figure 2 D, there was a significant ROI  $\times$  task  $\times$  response interaction. This interaction was because signal magnitudes in SMG were higher during vocal- (Re, Pr) than during button-response blocks in the discrimination but not in the 2-back tasks. This effect was not present in the STG ROIs (HG, aPT, pPT).

Our next analyses focused on motor-response type (Re, Bu), vocal-response type (Re, Pr), and vowel type (Ph, NPh) effects (Fig. 4). For these analyses, the ROI data were collapsed across discrimination and 2-back tasks as no significant task  $\times$  stimulus type  $\times$  response type interaction was detected in the omnibus ANOVA (because the stimulus type and response type main effects were observed in both tasks). The results of three- and two-way repeated-measures ANOVAs testing these effects in each ROI are shown in Table 5. A significant main effect of motor-response type (Re, Bu) was present in all ROIs (cf. Fig. 2 C). Further, a significant interaction between hemisphere and vocal-response type (Re, Pr) was observed in pPT and SMG during the vowel tasks (Re > Pr in the left hemisphere; cf. Fig. 3 A), whereas during the pitch tasks there was a vocal-response type main effect in HG and aPT (Pr > Re,

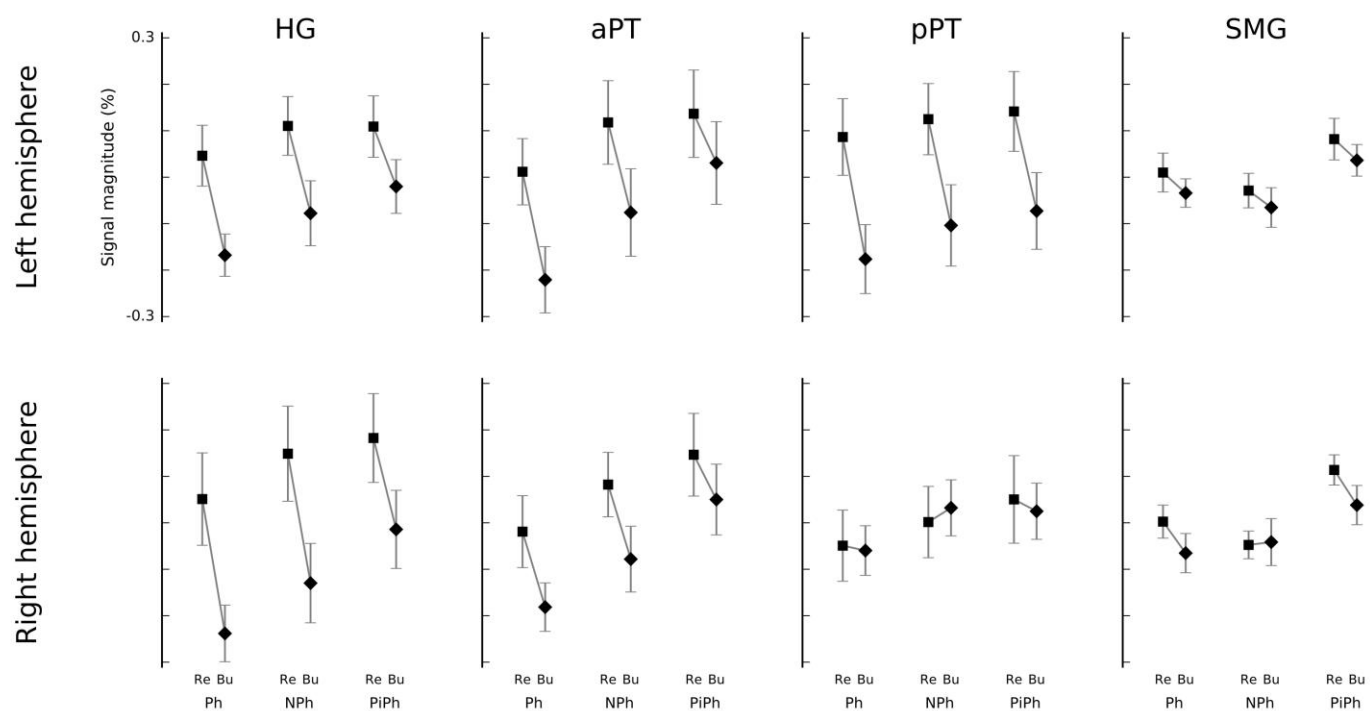


cf. Fig. 3 B). No significant interactions were observed between vowel type (Ph, NPh) and vocal-response type (Re, Pr).

The signal magnitudes in the count task (Fig. 5) were analyzed using repeated measures ANOVAs with factors hemisphere, stimulus (Ph, NPh, phPi), and motor-response type (Re, Bu). The main effect of motor-response type was significant in all STG ROIs and the hemisphere  $\times$  motor-response type interaction was significant in the pPT ROI (Table 6).



**Fig. 4.** ROI analysis of motor-response type effects during Ph, NPh and PiPh blocks (discrimination and 2-back tasks are collapsed together; full-block data; Table 4). Mean ( $\pm$ SEM; N = 20) relative signal magnitude in each ROI and condition is shown. To facilitate visual comparison, mean signal magnitude across all ROIs and conditions is subtracted from the values shown.



**Fig. 5.** ROI signal magnitudes in the control count task (details as in Fig. 4).

**Table 4.** The results of the omnibus five-way ANOVA (FDR corrected within the ANOVA) conducted on the ROI data with factors ROI (HG, aPT, pPT, SMG), hemisphere, task (discrimination, 2-back), stimulus (Ph, NPh, piPh), response (Re, Pr, Bu). The main effects of hemisphere and ROI and their interaction are not reported (in Tables 4–10) as the absolute signal magnitude difference between anatomical regions is not necessarily meaningful.

	<b>df</b>	<b>F</b>	<b>P</b>
<b>stimulus</b>	2,38	6.6	<b>0.007</b>
<b>response</b>	2,38	24	<b>0.001</b>
<b>ROI × task</b>	3,117	93	<b>0.001</b>
<b>ROI × response</b>	6,114	8.8	<b>0.001</b>
<b>hemisphere × stimulus</b>	2,38	5.5	<b>0.020</b>
<b>stimulus × response</b>	4,76	5.2	<b>0.003</b>
<b>ROI × hemisphere × response</b>	6,114	12	<b>0.001</b>
<b>ROI × task × response</b>	6,114	6.1	<b>0.001</b>
<b>ROI × stimulus × response</b>	12,228	5.6	<b>0.001</b>
task	1,19	0.6	0.543
ROI × stimulus	6,114	1.3	0.366
hemisphere × task	1,19	1.2	0.366
hemisphere × response	2,38	1.6	0.335
task × stimulus	2,38	1.5	0.369
task × response	2,38	0.4	0.723
ROI × hemisphere × task	3,117	3.3	0.055
ROI × hemisphere × stimulus	6,114	0.9	0.581
ROI × task × stimulus	6,114	0.4	0.853
hemisphere × task × stimulus	2,38	2.9	0.243
hemisphere × task × response	2,38	2.5	0.189
hemisphere × stimulus × response	4,76	1.3	0.367
task × stimulus × response	4,76	0.6	0.701
ROI × hemisphere × task × stimulus	6,114	2.5	0.058
ROI × hemisphere × task × response	6,114	1.3	0.373
ROI × hemisphere × stimulus × response	12,228	1.5	0.245
ROI × stimulus × task × response	12,228	2.0	0.055
hemisphere × task × stimulus × response	4,76	0.5	0.752
ROI × hemisphere × stimulus × task × response	12,228	0.9	0.577

**Table 5.** The results of ANOVAs conducted on the ROI data (discrimination and 2-back tasks collapsed). P-values are FDR corrected within and across ANOVAs (in Tables 5–10). For vowel tasks, the ANOVAs tested the effects of motor-response type (Re, Bu), vocal-response type (Re, Pr), and vowel type (Ph, NPh). For pitch tasks (piPh stimuli), the ANOVAs tested the effects of motor-response type (Re, Bu) and vocal-response type (Re, Pr). The ANOVAs were conducted separately for each ROI. Degrees of freedom are 1,19 in all cases.

	HG		aPT		pPT		SMG	
	F	P	F	P	F	P	F	P
<b>Vowel tasks: hemisphere × vowel type (Ph, NPh) × motor-response type (Re, Bu)</b>								
vowel type	13	<b>0.008</b>	7.4	<b>0.037</b>	0.2	0.753	21	<b>0.002</b>
motor-response type	63	<b>0.001</b>	65	<b>0.001</b>	34	<b>0.001</b>	10	<b>0.016</b>
hemisphere × motor-response type	13	<b>0.008</b>	0.3	0.734	10	<b>0.016</b>	30	<b>0.001</b>
hemisphere × vowel type	0.5	0.661	2.3	0.278	0.2	0.754	5.6	0.073
vowel type × motor-response type	1.9	0.343	1.1	0.503	0.3	0.728	0.7	0.605
hemisphere × vowel type × motor-response type	0.3	0.729	3.2	0.186	4.2	0.657	0.4	0.734
<b>Vowel tasks: hemisphere × vowel type (Ph, NPh) × vocal-response type (Re, Pr)</b>								
hemisphere × vocal-response type	0.2	0.759	2.4	0.272	18	<b>0.002</b>	9.0	<b>0.024</b>
vowel type	0.8	0.582	1.9	0.343	1.9	0.335	4.1	0.129
vocal-response type	4.1	0.126	6.0	0.059	4.6	0.102	6.3	0.053
hemisphere × vowel type	0.3	0.737	1.2	0.464	1.5	0.389	2.0	0.323
vowel type × vocal-response type	0.6	0.653	0	0.911	0.5	0.680	0.2	0.788
hemisphere × vowel type × vocal-response type	0.9	0.554	3.0	0.208	0.4	0.667	0.1	0.868
<b>Pitch tasks: hemisphere × motor-response type (Re, Bu)</b>								
hemisphere × motor-response type	10	<b>0.013</b>	2.9	0.234	1.0	0.502	20	<b>0.002</b>
motor-response type	0.4	0.728	0.2	0.776	2.2	0.304	3.9	0.143
<b>Pitch tasks: hemisphere × vocal-response type (Re, Pr)</b>								
vocal-response type	13	<b>0.008</b>	7.0	<b>0.042</b>	0.1	0.866	0.4	0.723
hemisphere × vocal-response type	1.7	0.352	0.2	0.791	0.1	0.909	0.7	0.609

**Table 6.** The results of ANOVAs conducted on the ROI data of the count task. The ANOVAs tested the effects of motor-response type (Re, Bu) and stimulus (Ph, NPh, piPh).

	HG			aPT		pPT		SMG	
	df	F	P	F	P	F	P	F	P
stimulus	2,38	12	<b>0.009</b>	20	<b>0.002</b>	5.2	<b>0.028</b>	7.5	<b>0.011</b>
motor-response type	1,19	55	<b>0.001</b>	33	<b>0.001</b>	7.9	<b>0.030</b>	6.9	<b>0.043</b>
hemisphere × motor-response type	1,19	5.1	0.078	2.5	0.251	35	<b>0.001</b>	0.9	0.350
hemisphere × stimulus	2,38	3.3	0.112	0.9	0.605	0.1	0.916	0.2	0.811
stimulus × motor-response type	2,38	1.5	0.391	1.7	0.335	0.3	0.831	2.7	0.081
hemisphere × stimulus × motor-response type	2,38	0.1	0.907	1.4	0.434	0.5	0.730	0.2	0.824

### 3.4 Analysis of the Ph and NPh vowels uttered during fMRI

Naïve listeners classified the utterances produced during fMRI as Ph or NPh vowels with a mean accuracy of 67 % correct (significantly above 50 % chance, permutation interference,  $P < 0.001$ ). The classification accuracy was higher than 54 % (significantly above chance, permutation test, FDR corrected  $P < 0.05$  in all cases) for (utterances produced by) each fMRI subject.

### 3.5 Additional ROI analyses

The results of the full-block (FB) and block-start analysis (BS) were compared with each other using the same ANOVAs (as in section 3.3) with an additional analysis (FB, BS) factor (Table 7). These comparisons were conducted, in particular, to test whether activation to self-produced vocalizations (not present during the block start) affected the main effect of motor-response type (Re, Bu) in the FB analysis. However, no analysis × motor-response type (Re, Bu) interactions were observed in any of the ROIs because the main effect of motor-response type observed in the FB analysis was significant also in the BS analysis in most of the ROIs (Table 8).

Incidentally, the ANOVAs with factors analysis (FB, BS), hemisphere, vowel type (Ph, NPh) and vocal-response type (Re, Pr) revealed one interaction of interest (i.e. involving the analysis and motor- or vocal response-type factors, Table 7). In the aPT ROI, the analysis  $\times$  vocal-response type interaction was significant because in the FB analysis activation was stronger during repetition than production responses, whereas in the BS analysis this relationship was reversed. A similar (nonsignificant) trend was observed in HG.

Analogous comparisons between the results of the first-response (FR) and FB analysis revealed no significant interactions of interest (Table 9). These comparisons were conducted, in particular, to test whether stimulus-dependent suppression effects (due to repetitions of the self-produced vowels in the production blocks) affected the main effect of vocal-response type (Re, Pr) in the FB analysis (Note that this analysis also controlled for the possible effects due to differences in the amount of vocalization responses). However, no analysis  $\times$  vocal-response type was observed as similar vocal-response type effects were observed both analyses (Table 10).

**Table 7.** The results of the comparisons conducted between the full-block (FB) and block-start analysis (BS). Significant interactions of interest (i.e interactions involving analysis and motor/vocal-response type) are listed first and bolded (in Table 7 and 9). Note that only one significant interaction of interest (i.e. analysis  $\times$  vocal-response in aPT) was found. Degrees of freedom are 1,19 in all cases (in Tables 7–10).

509

	HG		aPT		pPT		SMG	
	F	P	F	P	F	P	F	P
<b>Vowel tasks: analysis (BS, FB) <math>\times</math> hemisphere <math>\times</math> vowel type (Ph, NPh) <math>\times</math> motor-response type (Re, Bu)</b>								
analysis $\times$ vowel type	6.7	0.080	4.9	0.154	2.1	0.393	0.1	0.949
analysis $\times$ hemisphere $\times$ vowel type	0.5	0.765	0.5	0.733	0.9	0.648	2.2	0.386
analysis $\times$ motor-response type	4.5	0.168	3.9	0.198	1.3	0.541	0.6	0.707
analysis $\times$ hemisphere $\times$ motor-response type	0.6	0.707	0	0.955	1.9	0.420	0.9	0.645
analysis $\times$ vowel type $\times$ motor-response type	0.3	0.842	0.5	0.765	0.1	0.945	0	0.986
analysis $\times$ hemisphere $\times$ vowel type $\times$ motor-response type	1.4	0.525	1.0	0.626	0	0.951	0	0.991
vowel type	20	0.001	17	0.001	8.2	0.051	3.1	0.261
motor-response type	35	0.001	34	0.007	24	0.001	32	0.001
hemisphere $\times$ vowel type	1.7	0.459	1.7	0.459	1.9	0.420	8.8	0.044
hemisphere $\times$ motor-response type	4.8	0.151	0.3	0.843	3.3	0.248	11	0.026
vowel type $\times$ motor-response type	0.4	0.794	0.2	0.887	0	0.991	0.7	0.670
hemisphere $\times$ vowel type $\times$ motor-response type	1.8	0.420	3.9	0.198	5.3	0.135	0.1	0.925
<b>Vowel tasks: analysis (BS, FB) <math>\times</math> hemisphere <math>\times</math> vowel type (Ph, NPh) <math>\times</math> vocal-response type (Re, Pr)</b>								
<b>analysis <math>\times</math> vocal-response type</b>	5.2	0.136	10	<b>0.030</b>	3.8	0.206	1.2	0.570
analysis $\times$ vowel type	1.8	0.448	4.3	0.174	0	0.951	0.9	0.646
analysis $\times$ hemisphere $\times$ vowel type	1.1	0.587	0	0.951	0	0.952	2.9	0.284
analysis $\times$ hemisphere $\times$ vocal-response type	0.3	0.842	0	0.949	0.8	0.656	0.2	0.843
analysis $\times$ vowel type $\times$ vocal-response type	3.2	0.262	2.3	0.362	1.1	0.587	0.4	0.795
analysis $\times$ hemisphere $\times$ vowel type $\times$ vocal-response type	0	0.991	0.2	0.888	0	0.950	0	0.986
vowel type	4.7	0.154	10	0.026	1.4	0.504	0.2	0.843
vocal-response type	0.6	0.733	1.0	0.606	0.1	0.925	0.8	0.668
hemisphere $\times$ vowel type	2.2	0.388	0.5	0.753	0	0.986	8.6	0.044
hemisphere $\times$ vocal-response type	0	0.951	0.3	0.841	1.5	0.489	7.3	0.047
vowel type $\times$ vocal-response type	4.1	0.194	2.2	0.379	2.7	0.320	1.0	0.611
hemisphere $\times$ vowel type $\times$ vocal-response type	0.4	0.809	1.2	0.574	0.1	0.986	0	0.986
<b>Pitch tasks: analysis (BS, FB) <math>\times</math> hemisphere <math>\times</math> motor-response type (Re, Bu)</b>								
analysis $\times$ motor-response type	0.8	0.668	1.7	0.459	1.9	0.420	6.4	0.088
analysis $\times$ hemisphere $\times$ motor-response type	4.3	0.174	0.8	0.668	1.5	0.484	0	0.986
motor-response type	2.0	0.413	3.1	0.266	10	0.026	19	0.001
hemisphere $\times$ motor-response type	9.6	0.035	0.1	0.951	1.9	0.420	20	0.001
<b>Pitch tasks: analysis (BS, FB) <math>\times</math> hemisphere <math>\times</math> vocal-response type (Re, Pr)</b>								
analysis $\times$ vocal-response type	0.2	0.843	0.9	0.648	1.2	0.569	0.8	0.668
analysis $\times$ hemisphere $\times$ vocal-response type	0.3	0.816	1.1	0.600	0.1	0.993	0	0.986
vocal-response type	13	0.014	12	0.021	12	0.021	0.2	0.890
hemisphere $\times$ vocal-response type	0	0.978	0.2	0.843	0	0.986	0.5	0.765

510

511 **Table 8.** The results of ANOVAs testing effects in the BS analysis (using the same  
512 ANOVAs as in the FB analysis, Table 5).

513

	HG		aPT		pPT		SMG	
	F	P	F	P	F	P	F	P
<b>Vowel tasks: hemisphere × vowel type (Ph, NPh) × motor-response type (Re, Bu)</b>								
<b>motor-response type</b>	14	<b>0.010</b>	12	<b>0.024</b>	3.7	0.189	16	<b>0.010</b>
vowel type	6.1	0.096	3.7	0.187	5.2	0.118	0.6	0.687
hemisphere × vowel type	1.1	0.497	1.1	0.497	1.5	0.423	5.6	0.110
hemisphere × motor-response type	0.9	0.548	0.2	0.889	0.3	0.802	1.3	0.466
vowel type × motor-response type	0	0.964	0	0.964	0	0.964	0.3	0.802
hemisphere × vowel type × motor-response type	1.7	0.391	2.4	0.302	2.1	0.326	0	0.964
<b>Vowel tasks: hemisphere × vowel type (Ph, NPh) × vocal-response type (Re, Pr)</b>								
<b>vowel type</b>	3.6	0.191	11	<b>0.030</b>	0.6	0.687	0	0.964
vocal-response type	2.5	0.302	4.8	0.129	1.5	0.423	0	0.964
hemisphere × vowel type	2.1	0.326	0.2	0.865	0	0.964	6.3	0.094
hemisphere × vocal-response type	0.2	0.891	0.1	0.964	0	0.964	2.3	0.314
vowel type × vocal-response type	4.4	0.145	3.3	0.204	2.4	0.302	1.1	0.497
hemisphere × vowel type × vocal-response type	0.1	0.915	0.6	0.686	0.1	0.964	0	0.991
<b>Pitch tasks: hemisphere × motor-response type (Re, Bu)</b>								
<b>motor-response type</b>	2.2	0.326	5.3	0.118	9.2	<b>0.046</b>	19	<b>0.001</b>
hemisphere × motor-response type	0.2	0.889	0.1	0.916	4.7	0.129	4.7	0.129
<b>Pitch tasks: hemisphere × vocal-response type (Re, Pr)</b>								
vocal-response type	5.9	0.105	3.9	0.183	0.3	0.801	0.5	0.729
hemisphere × vocal-response type	0	0.964	0.1	0.915	0.2	0.860	0.1	0.891

514



**Table 9.** The results of the comparisons conducted between the full-block (FB) and first-response analysis (FR). Note that no significant interactions involving the analysis factor (listed first) were observed.

518

	HG		aPT		pPT		SMG	
	F	P	F	P	F	P	F	P
<b>Vowel tasks: analysis (FR, FB) × hemisphere × vowel type (Ph, NPh) × motor-response type (Re, Bu)</b>								
analysis × vowel type	0.3	0.843	1.0	0.606	0	0.955	3.6	0.229
analysis × hemisphere × vowel type	2.5	0.339	0.5	0.765	0.1	0.905	0.1	0.914
analysis × motor-response type	0	0.986	0.2	0.894	0.2	0.890	0.4	0.797
analysis × hemisphere × motor-response type	0.1	0.925	0.7	0.668	0.3	0.841	0.6	0.707
analysis × vowel type × motor-response type	0.1	0.925	0.1	0.925	0.2	0.843	1.1	0.600
analysis × hemisphere × vowel type × motor-response type	0.2	0.892	0.4	0.795	0.9	0.646	0.1	0.908
vowel type	5.2	0.136	10	0.030	0.3	0.843	2.7	0.316
motor-response type	70	0.001	66	0.001	31	0.001	8.6	0.049
hemisphere × vowel type	13	0.014	1.5	0.489	11	0.026	24	0.001
hemisphere × motor-response type	0.6	0.713	2.3	0.364	1.0	0.611	0.1	0.933
vowel type × motor-response type	2.3	0.362	0.4	0.796	0.1	0.926	0.5	0.766
hemisphere × vowel type × motor-response type	0	0.986	0.1	0.945	1.1	0.585	0.7	0.692
<b>Vowel tasks: analysis (FR, FB) × hemisphere × vowel type (Ph, NPh) × vocal-response type (Re, Pr)</b>								
analysis × vowel type	0.4	0.795	0.9	0.639	0	0.986	0.2	0.842
analysis × vocal-response type	0.8	0.668	0	0.952	0.2	0.843	0.1	0.914
analysis × hemisphere × vowel type	0	0.986	1.9	0.420	0.1	0.945	0.7	0.702
analysis × hemisphere × vocal-response type	1.3	0.556	2.0	0.407	0.1	0.921	0.6	0.707
analysis × vowel type × vocal-response type	0.5	0.734	0.1	0.925	0.5	0.759	0	0.986
analysis × hemisphere × vowel type × vocal-response type	0.6	0.732	0	0.985	0	0.986	0.4	0.794
vowel type	3.4	0.247	5.4	0.128	1.1	0.585	1.4	0.504
vocal-response type	0.8	0.663	2.9	0.293	7.3	0.066	6.9	0.076
hemisphere × vowel type	0.2	0.843	7.0	0.074	0.1	0.952	0.2	0.843
hemisphere × vocal-response type	3.3	0.247	0	0.952	8.9	0.044	9.8	0.035
vowel type × vocal-response type	0	0.986	0	0.979	0	0.986	0.2	0.894
hemisphere × vowel type × vocal-response type	2.5	0.349	1.7	0.452	0.2	0.896	0.2	0.843
<b>Pitch tasks: analysis (FR, FB) × hemisphere × motor-response type (Re, Bu)</b>								
analysis × hemisphere	0.1	0.966	0	0.966	4.3	0.186	5.6	0.117
analysis × motor-response type	4.8	0.152	4.8	0.152	0.3	0.817	4.0	0.198
analysis × hemisphere × motor-response type	0.1	0.945	0.1	0.945	3.1	0.266	0.6	0.713
hemisphere × motor-response type	8.6	0.048	7.3	0.066	1.6	0.482	19	0.001
motor-response type	0	0.951	0	0.951	1.7	0.459	2.6	0.325
<b>Pitch tasks: analysis (FR, FB) × hemisphere × vocal-response type (Re, Pr)</b>								
analysis × vocal-response type	0.9	0.648	0.8	0.663	0.1	0.908	2.4	0.346
analysis × hemisphere × vocal-response type	1.1	0.600	0.1	0.991	1.1	0.587	0.5	0.733
vocal-response type	12	0.021	8.2	0.049	0	0.986	3.5	0.242
hemisphere × vocal-response type	0.3	0.843	0.1	0.925	0.9	0.656	1.7	0.451

519

520 **Table 10.** The results of the FR analysis were analyzed using the same ANOVAs as  
521 in the FB analysis (Table 5).

522

	HG		aPT		pPT		SMG	
	F	P	F	P	F	P	F	P
<b>Vowel tasks: hemisphere × vowel type (Ph, NPh) × motor-response type (Re, Bu)</b>								
motor-response type	64	<b>0.001</b>	50	<b>0.001</b>	26	<b>0.001</b>	10	<b>0.017</b>
hemisphere × motor-response type	10	<b>0.036</b>	1.2	0.552	9.0	<b>0.037</b>	36	<b>0.001</b>
vowel type	1.1	0.552	3.4	0.245	0	0.978	0.2	0.928
hemisphere × vowel type	0.4	0.887	1.6	0.455	0.7	0.720	0.2	0.930
vowel type × motor-response type	1.8	0.431	0.1	0.957	0	0.978	0	0.958
hemisphere × vowel type × motor-response type	0.3	0.960	0	0.960	2.0	0.404	0.2	0.929
<b>Vowel tasks: hemisphere × vowel type (Ph, NPh) × vocal-response type (Re, Pr)</b>								
vocal-response type	0	0.960	0.8	0.671	4.0	0.187	8.2	<b>0.048</b>
hemisphere × vocal-response type	2.3	0.365	1.1	0.552	9.1	<b>0.037</b>	0	0.978
vowel type	1.9	0.431	4.0	0.187	0.3	0.877	0.2	0.928
hemisphere × vowel type	0	0.960	4.2	0.188	0.1	0.957	0	0.978
vowel type × vocal-response type	0.2	0.929	0.1	0.958	0.2	0.929	0	0.978
hemisphere × vowel type × vocal-response type	1.7	0.454	0.4	0.851	0	0.960	0.6	0.744
<b>Pitch tasks: hemisphere × motor-response type (Re, Bu)</b>								
hemisphere × motor-response type	2.9	0.284	0	0.978	2.8	0.294	16	<b>0.014</b>
motor-response type	0.1	0.958	0	0.978	1.1	0.563	1.5	0.464
<b>Pitch tasks: hemisphere × vocal-response type (Re, Pr)</b>								
vocal-response type	5.69	0.112	7.2	0.067	0.1	0.958	4.0	0.187
hemisphere × vocal-response type	0.64	0.720	0	0.963	1.1	0.552	1.5	0.464

523

## 4 Discussion

The interaction of auditory-sensory processing and motor actions plays an important role in current theoretical work on AC (Hickok, 2016; Hickok and Poeppel, 2007; Rauschecker, 2010; Rauschecker and Romanski, 2011; Rauschecker and Scott, 2009). During auditory-motor tasks, signals from motor cortex suppress responses to self-produced vocalizations (Agnew et al., 2013; Eliades and Wang, 2003) and motor programs are fine-tuned based on auditory feedback (Purcell and Munhall, 2006; Tachibana et al., 2010; Tourville et al., 2008). Motor execution may modulate responses in AC even when the motor task is not directly associated with sound production (Schneider et al., 2014; Wikman et al., 2015). Further, a series of human imaging studies has highlighted the role of left posterior STG in auditory-motor integration during vocalization (Alho et al., 2012; Hickok et al., 2011; Hickok et al., 2009; Okada et al., 2010; Peschke et al., 2012; Peschke et al., 2009; Shuster and Lemieux, 2005; Simmonds et al., 2014b). The present study investigated whether auditory-motor effects interact with the strong task-dependent modulations in AC and adjacent regions (Rinne et al., 2009; Harinen et al., 2013; Harinen and Rinne, 2014; Häkkinen et al., 2015). Our subjects performed demanding discrimination and 2-back tasks on vowels during fMRI and responded to targets either by repeating the last vowel of the target pair or humming its pitch, producing a given response vowel, or pressing a response button. In line with our previous results, activation in wide STG/IPL regions was strongly modulated by task (discrimination vs. 2-back). As novel findings, we found that (1) activation in these regions also depended on the motor-response type, i.e. whether subjects performed the tasks using vocal or button responses, (2) the task and motor-response type effects interacted in IPL but not in

STG, and (3) left-hemisphere regions extending from mid STG to IPL showed stronger activation during vowel-repetition (vocalization based on direct auditory-to-motor translation) than during vowel-production (vocalization not based on auditory input) blocks.

#### *4.1 Main effects of task and motor-response type*

The task-dependent modulations shown in Figure 2 B were due to stronger activation in STG during the discrimination than 2-back memory task, whereas IPL showed stronger activation during the 2-back memory task. Based on our previous studies using analogous tasks, it is clear that both tasks were associated with enhanced activation in STG (see, e.g. Fig. 3 B of Harinen and Rinne, 2013). However, in the present study, we were not interested in the general effects of auditory tasks and, therefore, a baseline condition (e.g. the same vowels presented during a demanding visual task) to extract these effects was not included in the design. It is also important to note that the present task-dependent modulations are not specific to vowel processing, as similar effects were observed in our previous studies using analogous discrimination and 2-back tasks performed on sounds varying in pitch or location (Häkkinen et al., 2015; Rinne et al., 2009; Rinne et al., 2012). Moreover, our previous results suggest that the activation enhancements in IPL during 2-back tasks are related to general operations on categorical representations, whereas the enhanced STG activation during discrimination tasks are due to analysis of the acoustical features of the sounds (Harinen and Rinne, 2014).

In the present study, most of the regions showing task-dependent effects in STG and IPL were also modulated by motor-response type (Fig. 2 C). This was due to stronger activation in extensive regions of AC and IPL when subjects responded to

targets by (overtly) uttering a vowel (repetition and production responses) than when they performed identical tasks but responded with button presses (Fig. 4). To our knowledge, the present study is the first one to compare activation in STG–IPL region during vocal and manual responding during active listening tasks. What might be the cause of the enhanced activation associated with vocal responding?

First, although stimulus-dependent activation to vowels is typically observed in bilateral regions in or near HG (see, e.g., Fig. 3 A of Harinen and Rinne, 2013), activation elicited by the self-produced vowels could extend to more posterior STG regions (Huang et al., 2002). Thus, the enhanced activation during vocal responding could be due to effects related to the self-produced vocalizations. To test this possibility, we separately analyzed the activation associated with the first seconds of each task block (from block onset to the first response). We reasoned that if the enhanced activation during vocal-response blocks is due to the self-vocalized vowels, then the activation difference between vocal- and button-response blocks should not be present during the beginning of the blocks when subjects are engaged in the tasks but have not yet responded to targets. However, our analyses revealed no significant analysis  $\times$  motor-response type interaction effects as activation was stronger during vocal- than button-response blocks also during the beginning of the task blocks (Tables 7 and 8). Thus, the enhanced activation during vocal responses cannot be explained by stimulus-dependent activation to the self-produced vowels or by any effects related to the actual execution of the motor responses.

Second, it could also be argued that the enhanced activation during vocal-response blocks is due to increased general task difficulty as performance was systematically lower during vocal than button responding (probably because vowel responding interfered with performance in the vowel tasks; Table 2). However, this is

unlikely as activation was stronger during vocal than button responding also in the control count task (Fig. 5), where performance did not significantly differ between vocal- and button-response blocks (Tables 2 and 3, note that vocal- and button-response RTs are not directly comparable). Further, the results of our previous studies show that increased general task difficulty during discrimination and n-back tasks, as such, is not associated with a uniform activation increase in STG–IPL region. For example, Rinne et al. (2009) reported that, during pitch discrimination, increasing task difficulty did not significantly modulate STG activation. By contrast, during n-back tasks, increasing task difficulty resulted in decreased STG activation and enhanced IPL activation (for similar results during discrimination and n-back tasks, see Rinne et al., 2012; Häkkinen et al., 2015; Harinen et al., 2013). Taken together, these results suggest that the enhanced activation during vocal-response blocks is not due to general task difficulty but is likely to be related to specific task requirements (see points 3–5 below) during vocal and button responding.

Third, the enhanced activation during vocal responding could be because vocal responding required more mental rehearsal of vocalizations than button responding. Previous studies have shown that auditory imagery activates STG regions (Linke and Cusack, 2015; Zvyagintsev et al., 2013). It has also been suggested that auditory imagery contributes to activation in posterior STG regions during both covert and overt speech production (Parker Jones et al., 2014). Auditory imagery would also explain the result that stronger activation during vocal than manual responding was observed already during the beginning of the block (before responses) as subjects could have used imagery to prepare for vocal responding.

Fourth, regions involved in auditory-motor integration could show stronger activation during vocal- than button-response blocks, and these regions could

be activated in a preparatory manner already in the beginning of the task blocks before any responses are executed. However, the results of previous studies (Buchsbaum et al., 2011; Hickok et al., 2003a; Hickok et al., 2009) predict that auditory-motor integration is associated with increased activation particularly in left posterior STG, whereas in the present study the difference between vocal- and button-response blocks was observed in wide STG and IPL regions bilaterally.

Fifth, it is also possible that the activation difference between vocal- and button-response blocks is due to stronger suppression of activation during manual (button) than during vocal responding. Previous studies have shown that both overt vocalizations and manual responses are associated with motor suppression effects (Heinks-Malanado et al., 2005; Wikman et al., 2015). In the present study, motor suppression effects could have been smaller in vocal- than button response blocks, for example, because the sensory consequences of the motor acts were highly relevant and prominent during vocal responding but irrelevant and nonexistent for button responses. Thus, according to this account, the activation difference between vocal- and button-response blocks was caused by signals from the motor cortex modulating processing in wide STG and IPL regions. This account would also explain the effects in the beginning of the blocks as auditory processing is modulated already during movement intention (Timm et al., 2014).

Interestingly, the comparison of signal magnitudes in the block-start and full-block analysis revealed that in the beginning of the blocks signal magnitudes were weaker during repetition- than production-response blocks, whereas in the full-block analysis the opposite pattern was observed (significant analysis  $\times$  vocal-response type interaction in the aPT). This incidental finding (based on less data than the full block analysis) could be related to the fact that in the production-response

blocks, the response vowel was given before the beginning of the block and, thus, subjects were able to mentally rehearse or prepare to vocalize it even before the first target. By contrast, in the repetition blocks, the vocalization responses depended on the target and the response vowel was therefore not known in the beginning of a block. Therefore, this result suggests that mental rehearsal of vocalizations significantly contributed to the activation during vocal-response blocks. However, mental rehearsal alone cannot fully explain the activation difference between vocal- and button-response blocks, as enhanced activation was observed also in the beginning of repetition-response blocks, where the vowel-to-be-uttered was not known and, thus, could not be mentally rehearsed. Taken together, it is likely that the present motor-response type effect is due to a combination of factors related to motor execution (e.g., suppression of activation during button-response blocks) and vocal responding (e.g., mental rehearsal of the vowel responses).

#### *4.2 Interaction of task and auditory-motor effects (H3)*

An interaction of task and motor-response type effects was observed in bilateral IPL but not in STG (Fig. 2 D, Table 4; no significant task  $\times$  vocal-response type or task  $\times$  vowel type interactions). Previous literature has implicated IPL in working memory (Koelsch et al., 2009; Leung and Alain, 2010; Gaab et al., 2006) and categorical processing (Harinen and Rinne, 2014; Husain et al., 2006; Raizada and Poldrack, 2007). In the present study, the interaction in IPL could be because the discrimination task required more resources for working memory (which vocalization to make) and/or categorical processing (maintenance of vowel categories) during vocal than during button responding, whereas in the (categorical) 2-back task working memory and categorical processing was an essential component of the task with both response



types. The lack of interaction effects in STG regions, in turn, suggests that task and motor effects in these regions are caused by independent mechanisms.

#### *4.3 Vowel repetition vs. vowel production (H1)*

Based on previous literature (e.g. Parker Jones et al., 2014; Simmonds et al., 2014a; Simmonds et al., 2014b; Simmonds et al., 2011), we hypothesized that vowel-repetition responses requiring direct auditory-motor translation (repeat the last part of the just-heard target vowel) would place a higher load on auditory-motor integration than vowel-production responses (produce the same given response vowel to each target in a block). Consistently, we found stronger activation in left hemisphere regions extending from mid STG to IPL during repetition- than production-response blocks (Fig. 3 A).

It could also be argued that the stronger activation during repetition- than production-response blocks is due to stimulus-dependent suppression effects associated with the self-produced vowels. As the magnitude of the fMRI signal decreases with repetition of similar stimuli (Bergerbest et al., 2004), the activation to the vocalizations during production-response blocks, where the same response vowel was repeatedly uttered, is likely to be smaller than that during repetition-response blocks, where the uttered vowel was almost always different than the previous one. To investigate whether such adaptation contributed to the activation difference between repetition- and production-response blocks, we separately analyzed the activation elicited by the first vowel responses of each task block. The first vowel responses in each task block were preceded by at least a 16 s period (rest + block start) with no responses. Thus, in terms of adaptation effects, the activation to the first responses of repetition and production blocks should be identical. Yet, there were no significant

differences between the signal magnitudes associated with the first-responses of each block and those in the full-block analysis and similar vocal-response type effects were observed during both the full-block and first-response analysis (Tables 9 and 10).

Thus, adaptation of stimulus-dependent activation to the overt vocalizations does not explain the observed activation difference between repetition and production blocks.

It could also be argued that, in the production blocks, subjects covertly rehearsed the response vowel to-be-uttered and that this mental rehearsal adapted the representations for this vowel so that activation to the first vowel utterance (and the ones following that) were adapted in the production but not in the repetition block (where the response vowel was known only after the target). This account, however, seems unlikely since previous work suggests that neither mental imagery nor covert production are associated with strong stimulus-specific adaption effects (Tian and Poeppel, 2013). Further, it could be argued that covert rehearsal of response vowels in the production blocks is associated with suppression due to auditory-motor prediction (Eliades & Wang, 2003, 2005; Houde et al., 2002; Tian and Poeppel, 2015).

Auditory-motor prediction related suppression is observed in studies in which auditory feedback (i.e. the vocalization that was rehearsed) is presented immediately (<200 ms; Tian and Poeppel, 2015) after rehearsal. In the present study, subjects produced and heard their own utterances similarly in the production and repetition blocks. If subjects covertly rehearsed the response vowel in the production blocks, then this occurred well before (>200 ms) the production of the actual vowel response. Thus, the stronger activation during repetition- than production-response blocks cannot be easily explained by mental rehearsal of response vowels in the production blocks.

As noted in section 4.1, activation during the beginning of a block (before any

responses were made; block-start analysis) was weaker during repetition than production responses in STG ROIs (significant analysis  $\times$  vocal-response type interaction in the aPT). That is, stronger activation in the repetition response blocks emerged only at the time when the first vocal response of each task block was given and when auditory-to-motor translation was required. Taken together, these activation patterns observed during repetition- and production-response blocks support the notion that regions in left posterior STG and IPL are involved in auditory-motor integration during speech.

#### *4.4 Ph vs. NPh vowels (H2)*

The Ph and NPh vowels used in the present study were spectrotemporally quite similar to each other (Fig. 1 A). Thus, we assumed that the Ph and NPh vowels would not be associated with systematic stimulus-dependent activation differences. In a previous study (Harinen and Rinne, 2013), we presented the same Ph and NPh vowels during a demanding visual task to estimate stimulus-dependent effects (in the absence of directed auditory attention or task) to these vowels. In that study, we found stronger stimulus-dependent activation to Ph vowels in a few scattered clusters in STG and IPL regions but no enhanced activation to NPh vowels was observed. Yet, when the vowels were presented during auditory tasks (with button responses), in both the previous and the present study, activation in STG was stronger during task blocks with NPh than Ph vowels (see also Harinen and Rinne, 2014). In the previous study, we argued that this difference is because a more thorough spectrotemporal analysis is required to complete the tasks performed on NPh vowels, as tasks performed on Ph vowels benefit from speech-specific processing. Together, the present and our previous results indicate that activation in STG is sensitive to the speech-level

749 difference between Ph and NPh vowels during active listening.

750         In the present study, we hypothesized that the load on auditory-motor  
751 integration would be higher during repetition of NPh than Ph vowels, as repetition of  
752 Ph vowels could utilize the well-learned representations of native language, whereas  
753 repetition of NPh vowels would rely more on a direct translation of auditory input to  
754 motor commands. However, we found no systematic activation differences between  
755 NPh and Ph repetition-response blocks. In particular left posterior STG regions,  
756 where activation was stronger during repetition than production responses, showed no  
757 significant activation differences when subjects responded by repeating Ph or NPh  
758 vowels.

759         This pattern of results would be observed if subjects were not able to  
760 perceive and produce the difference between Ph and NPh vowels so that they uttered a  
761 Ph vowel also when they were supposed to repeat an NPh vowel. Obviously, if this is  
762 the case, then the requirements for audiomotor integration would have been identical  
763 in Ph and NPh blocks. To investigate this possibility, we presented the vowels uttered  
764 and recorded during fMRI to a group of naïve listeners. We asked the listeners to  
765 classify each utterance as a Finnish or non-Finnish vowel. Although the sound quality  
766 of the recorded vowel utterances was relatively poor, the naïve listeners were able to  
767 classify the Ph and NPh vowel utterances significantly above chance level (67 %  
768 accuracy rate on average). This indicates that subjects systematically perceived and  
769 reproduced the differences between Ph and NPh vowels during fMRI.

770         The lack of significant activation differences in left posterior STG  
771 between repetition of Ph and NPh vowels is in line with the results of a previous study  
772 in which activation in posterior STG did not significantly differ when subjects  
773 repeated words or pseudowords (Parker Jones et al., 2014). It could be argued,

however, that these comparisons (i.e., repetition of Ph vs. NPh vowels or words vs. pseudowords) are not strong tests for auditory-motor integration effects as auditory-to-motor translation is required in both cases. The present contrast between repetition and production blocks (discussed in section 4.3.) may be a stronger test as auditory-to-motor translation is required during repetition but not during production responses. Nevertheless, together the present and previous results suggest that activation in left posterior STG (and in other STG areas) does not strongly depend on whether subjects repeat native language or novel speech material. Thus, speech-level information, although clearly present at the level of AC, does not seem to strongly affect auditory-motor integration in posterior STG.

#### *4.5 Pitch repetition vs. vowel production*

Previous studies have shown that, in addition to listening to speech and covert speech production, left posterior STG regions are activated also during non-speech audiomotor tasks such as humming of melodies or playing a musical instrument (Buchsbaum et al., 2001; Hickok et al., 2003a, b; Hickok et al., 2009; Pa and Hickok, 2008). Based on the results of these studies, we expected that, similar to vowel repetition, also pitch repetition (humming the pitch of a target) would show enhanced activation relative to vowel-production responding in posterior STG. However, the comparisons between pitch-repetition and vowel-production response blocks during pitch tasks did not reveal enhanced activation associated with pitch repetition. In contrast, these comparisons showed enhanced activation in bilateral regions in and near HG during vowel-production responding (Fig. 3 and Fig. 4). It is unlikely that this activation difference is due to enhanced stimulus-dependent activation to self-vocalized vowels vs. humming as activation tended to be stronger during production

than repetition response blocks also in the beginning of the blocks when subjects were engaged in the task but had not yet responded to targets (Tables 7 and 8). However, it is possible that the specific effects related to pitch repetition were masked by some other differences between humming and vowel responding. For example, humming could be associated with stronger motor-suppression effects (see section 4.1) than vowel production. Thus, in future studies, pitch-repetition effects should be investigated using pitch-repetition and pitch-production responses with identical motor requirements (i.e., humming). Further, repetition of pitch intervals could be associated with stronger effects.

#### *4.6 Implications for auditory-motor integration*

It has been suggested that an area “Spt” (Sylvian parietal-temporal) acts as an interface between auditory and motor systems during speech and music production tasks. Spt is defined as an area in left posterior STG where activation increases during both perception and (covert) production of speech (Buchsbaum et al., 2001; Hickok, 2009; Hickok et al., 2003a; Pa and Hickok, 2008). It has also been argued that Spt is relatively more selective for vocal than manually mediated auditory-motor integration as it shows stronger activation when subjects covertly hum a melody (vocal effectors) than when they covertly play it on a piano (manual effectors; Hickok et al., 2003a). Although the present study did not focus on area Spt, our results are consistent with the general view that left posterior STG is involved in auditory-motor integration. In particular, the role of this region in auditory-to-motor translation is supported by the stronger activation during vowel repetition than production responses. Our results are also consistent with the idea that area Spt is relatively more selective for vocal than manual actions as we found stronger activation in left posterior STG during vocal

than button responding. However, in the present study, this effect was not specific to left posterior STG as vocal responding was associated with enhanced activation in wide STG and IPL regions bilaterally. This suggests that the difference between vocal and manual responding is a more general motor effect and that it is not specifically related to auditory-motor integration in posterior STG as such.

We found that activation in posterior STG did not depend on whether subjects repeated Ph or NPh vowels suggesting that the language-level distinction between Ph and NPh vowels does not affect auditory-to-motor translation in posterior STG and that this translation is conducted based on pure sensory acoustical information. This seems surprising as the results of the present and our previous studies show that language-level categorical vowel representations are available in STG during active listening tasks (Harinen et al., 2013; Harinen and Rinne, 2014) and as regions in left posterior STG are strongly implicated in speech-specific processing (Hickok, 2016; McGettigan et al., 2010; Zhang et al., 2011). Subsequent studies should aim to better understand the role of native language representations in the operations of left posterior STG.

The present results show that activation in left posterior STG during vocalization also depends on other characteristics of the task at hand. Further, task-dependent activation patterns drastically change within a few millimeters between posterior STG (showing enhanced activation during active listening) and IPL (decreased activation during discrimination, enhanced activation during 2-back). Attentional and task-dependent lability should, therefore, be carefully taken into account in studies investigating activation in these regions (e.g., definition of ROIs).

## 5 Conclusions

The present results show that activation in AC and adjacent regions during active audiomotor tasks is strongly modulated by the requirements of the task (discrimination vs. 2-back), motor-response type (vocal vs. button) and vocal-response type (repetition vs. production). Even seemingly small differences between the present experimental conditions (e.g., *prepare* to respond to targets by a vocal vs. button response) resulted in strong activation differences in STG and IPL regions. Importantly, the present study suggests that the task, motor-response type and vocal-response type effects are caused by independent mechanisms in AC. Future studies should aim to determine how these effects are controlled and mediated at the level of neuroanatomical pathways between AC and other brain regions.



## References

- Agnew, Z.K., McGettigan, C., Banks, B., Scott, S.K., 2013. Articulatory movements modulate auditory responses to speech. *Neuroimage* 73, 191–199.
- Alho, J., Sato, M., Sams, M., Schwartz, J.-L., Tiitinen, H., Jääskeläinen, I.P., 2012. Enhanced early-latency electromagnetic activity in the left premotor cortex is associated with successful phonetic categorization. *Neuroimage* 60, 1937–1946.
- Alho, K., Rinne, T., Herron, T.J., Woods, D.L., 2014. Stimulus-dependent activations and attention-related modulations in the auditory cortex: A meta-analysis of fMRI studies. *Hearing Research* 307, 29–41.
- Baldo, J.V., Klostermann, E.C., Dronkers, N.F., 2008. It's either a cook or a baker: Patients with conduction aphasia get the gist but lose the trace. *Brain and Language* 105, 134–140.
- Bergerbest, D., Ghahremani, D.G., Gabrieli, J.D.E., 2004. Neural correlates of auditory repetition priming: reduced fMRI activation in the auditory cortex. *Journal of Cognitive Neuroscience* 16, 966–977.
- Buchsbaum, B.R., Baldo, J., Okada, K., Berman, K.F., Dronkers, N., D'Esposito, M., Hickok, G., 2011. Conduction aphasia, sensory-motor integration, and phonological short-term memory – An aggregate analysis of lesion and fMRI data. *Brain and Language* 119, 119–128.
- Buchsbaum, B.R., Hickok, G., Humphries, C., 2001. Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science* 25, 663–678.
- Burnett, T.A., Freedland, M.B., Larson, C.R., Hain, T.C., 1998. Voice F0 responses to manipulations in pitch feedback. *Journal of the Acoustical Society of America* 103, 3153–3161.
- Eliades, S.J., Wang, X.Q., 2003. Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. *Journal of Neurophysiology* 89, 2194–2207.
- Gaab, N., Gaser, C., Schlaug, G., 2006. Improvement-related functional plasticity following pitch memory training. *Neuroimage* 31, 255–263.
- Golfinopoulos, E., Tourville, J.A., Bohland, J.W., Ghosh, S.S., Nieto-Castanon, A., Guenther, F.H., 2011. fMRI investigation of unexpected somatosensory feedback perturbation during speech. *Neuroimage* 55, 1324–1338.
- Greenlee, J.D.W., Jackson, A.W., Chen, F., Larson, C.R., Oya, H., Kawasaki, H., Chen, H., Howard, M.A., III, 2011. Human Auditory Cortical Activation during Self-Vocalization. *PLoS ONE* 6, 1–9.
- Guenther, F.H., 2006. Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders* 39, 350–365.

897 Guenther, F.H., Ghosh, S.S., Tourville, J.A., 2006. Neural modeling and imaging of  
898 the cortical interactions underlying syllable production. *Brain and Language* 96, 280–  
899 301.

900 Guenther, F.H., Vladusich, T., 2012. A neural theory of speech acquisition and  
901 production. *Journal of Neurolinguistics* 25, 408–422.

902 Hall, D.A., Haggard, M.P., Akeroyd, M.A., Summerfield, A.Q., Palmer, A.R., Elliott,  
903 M.R., Bowtell, R.W., 2000. Modulation and task effects in auditory processing  
904 measured using fMRI. *Human Brain Mapping* 10, 107–119.

905 Harinen, K., Aaltonen, O., Salo, E., Salonen, O., Rinne, T., 2013. Task-dependent  
906 activations of human auditory cortex to prototypical and nonprototypical vowels.  
907 *Human Brain Mapping* 34, 1272–1281.

908 Harinen, K., Rinne, T., 2013. Activations of human auditory cortex to phonemic and  
909 nonphonemic vowels during discrimination and memory tasks. *Neuroimage* 77, 279–  
910 287.

911 Harinen, K., Rinne, T., 2014. Acoustical and categorical tasks differently modulate  
912 activations of human auditory cortex to vowels. *Brain and Language* 138, 71–79.

913 Heinks-Maldonado, T.H., Mathalon, D.H., Gray, M., Ford, J.M., 2005. Fine-tuning of  
914 auditory cortex during speech production. *Psychophysiology* 42, 180–190.

915 Hickok, G., 2009. The functional neuroanatomy of language. *Physics of Life Reviews*  
916 6, 121–143.

917 Hickok, G., 2012. Computational neuroanatomy of speech production. *Nature*  
918 *Reviews Neuroscience* 13, 135–145.

919 Hickok, G., 2016. A cortical circuit for voluntary laryngeal control: Implications for  
920 the evolution language. *Psychonomic Bulletin and Review* 1, 56–63 .

921 Hickok, G., Buchsbaum, B., Humphries, C., Muftuler, T., 2003a. Auditory-motor  
922 interaction revealed by fMRI: Speech, music, and working memory in area Spt.  
923 *Journal of Cognitive Neuroscience* 15, 673–682.

924 Hickok, G., Buchsbaum, B., Humphries, C., Muftuler, T., 2003b. Auditory-motor  
925 interaction revealed by fMRI: Speech, music, and working memory in area Spt.  
926 *Journal of Cognitive Neuroscience* 15, 673–682.

927 Hickok, G., Houde, J., Rong, F., 2011. Sensorimotor integration in speech processing:  
928 computational basis and neural organization. *Neuron* 69, 407–422.

929 Hickok, G., Okada, K., Serences, J.T., 2009. Area Spt in the Human Planum  
930 Temporale Supports Sensory-Motor Integration for Speech Processing. *Journal of*  
931 *Neurophysiology* 101, 2725–2732.

932 Hickok, G., Saberi, K., 2012. Redefining the Functional Organization of the Planum  
933 Temporale Region: Space, Objects, and Sensory–Motor Integration, in Poeppel, D.,

- 934 Overath, T., Popper, A.N., Fay, R.R. (Eds.), *The Human Auditory Cortex*, Springer,  
935 p. 333–350.
- 936 Hickok, G., Saberi, D., 2007. The cortical organization of speech processing. *Nature*  
937 *Reviews Neuroscience* 8, 393–402.
- 938 Huang, J., Carr, T.H., Cao, Y., 2002. Comparing cortical activations for silent and  
939 overt speech using event-related fMRI. *Human Brain Mapping* 15, 39–53.
- 940 Husain, F.T., Fromm, S.J., Pursley, R.H., Hosey, L.A., Braun, A.R., Horwitz, B.,  
941 2006. Neural bases of categorization of simple speech and nonspeech sounds. *Human*  
942 *Brain Mapping* 27, 636–651.
- 943 Häkkinen, S., Ovaska, N., Rinne, T., 2015. Processing of pitch and location in human  
944 auditory cortex during visual and auditory tasks. *Frontiers in Psychology* 6, 1678.
- 945 Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Müller, K., Gruber, O., 2009.  
946 Functional architecture of verbal and tonal working memory: an FMRI study. *Human*  
947 *Brain Mapping* 30, 859–873.
- 948 Leung, A.W.S., Alain, C., 2010. Working memory load modulates the auditory 'What'  
949 and 'Where' neural networks. *Neuroimage* 55, 1260–1269.
- 950 Linke, A.C., Cusack, R., 2015. Flexible Information Coding in Human Auditory  
951 Cortex during Perception, Imagery, and STM of Complex Sounds. *Journal of*  
952 *Cognitive Neuroscience* 27, 1322–1333.
- 953 McGettigan, C., Warren, J.E., Eisner, F., Marshall, C.R., Shanmugalingam, P., Scott,  
954 S.K., 2010. Neural Correlates of Sublexical Processing in Phonological Working  
955 Memory. *Journal of Cognitive Neuroscience* 23, 673–682.
- 956 Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.H., Saberi, K., Serences, J.T.,  
957 Hickok, G., 2010. Hierarchical organization of human auditory cortex: evidence from  
958 acoustic invariance in the response to intelligible speech. *Cerebral Cortex* 20, 2486–  
959 2495.
- 960 Pa, J., Hickok, G., 2008. A parietal-temporal sensory-motor integration area for the  
961 human vocal tract: Evidence from an fMRI study of skilled musicians.  
962 *Neuropsychologia* 46, 362–368.
- 963 Parker Jones, O.P., Prejawa, S., Hope, T.M.H., Oberhuber, M., Seghier, M.L., Leff,  
964 A.P., Green, D.W., Price, C.J., 2014. Sensory-to-motor integration during auditory  
965 repetition: a combined fMRI and lesion study. *Frontiers in Human Neuroscience* 8.
- 966 Peschke, C., Ziegler, W., Eisenberger, J., Baumgaertner, A., 2012. Phonological  
967 manipulation between speech perception and production activates a parieto-frontal  
968 circuit. *Neuroimage* 59, 788–799.
- 969 Peschke, C., Ziegler, W., Kappes, J., Baumgaertner, A., 2009. Auditory-motor  
970 integration during fast repetition: The neuronal correlates of shadowing. *Neuroimage*  
971 47, 392–402.

- 972 Petkov, C.I., Kang, X., Alho, K., Bertrand, O., Yund, E.W., Woods, D.L., 2004.  
973 Attentional modulation of human auditory cortex. *Nature Neuroscience* 7, 658–663.
- 974 Prosek, R.A., Montgomery, A.A., Walden, B.E., Schwartz, D.M., 1979. Reaction-  
975 Time Measures of Stutterers and Nonstutterers. *Journal of Fluency Disorders* 4, 269–  
976 278.
- 977 Purcell, D.W., Munhall, K.G., 2006. Compensation following real-time manipulation  
978 of formants in isolated vowels. *Journal of the Acoustical Society of America* 119,  
979 2288–2297.
- 980 Raizada, R.D., Poldrack, R.A., 2007. Selective amplification of stimulus differences  
981 during categorical processing of speech. *Neuron* 56, 726–740.
- 982 Rauschecker, J.P., 2010. An expanded role for the dorsal auditory pathway in  
983 sensorimotor control and integration. *Hearing Research* 271, 16–25.
- 984 Rauschecker, J.P., Romanski, L.M., 2011. Auditory Cortical Organization: Evidence  
985 for Functional Streams. *The Auditory Cortex*, 99–116.
- 986 Rauschecker, J.P., Scott, S.K., 2009. Maps and streams in the auditory cortex:  
987 nonhuman primates illuminate human speech processing. *Nature Neuroscience* 12,  
988 718–724.
- 989 Rinne, T., 2010. Activations of Human Auditory Cortex During Visual and Auditory  
990 Selective Attention Tasks with Varying Difficulty. *Open Neuroimaging Journal* 4,  
991 187–193.
- 992 Rinne, T., Koistinen, S., Salonen, O., Alho, K., 2009. Task-dependent activations of  
993 human auditory cortex during pitch discrimination and pitch memory tasks. *The*  
994 *Journal of Neuroscience* 29, 13338.
- 995 Rinne, T., Koistinen, S., Talja, S., Wikman, P., Salonen, O., 2012. Task-dependent  
996 activations of human auditory cortex during spatial discrimination and spatial  
997 memory tasks. *Neuroimage* 59, 4126–4131.
- 998 Rinne, T., Muers, R.S., Salo, E., Slater, H., Petkov, C.I., 2017. Functional Imaging of  
999 Audio-Visual Selective Attention in Monkeys and Humans: How do Lapses in  
1000 Monkey Performance Affect Cross-Species Correspondences? *Cerebral Cortex*, 1–14.
- 1001 Rinne, T., Pekkola, J., Degerman, A., Autti, T., Jääskeläinen, I.P., Sams, M., Alho,  
1002 K., 2005. Modulation of auditory cortex activation by sound presentation rate and  
1003 attention. *Human Brain Mapping* 26, 94–99.
- 1004 Rogalsky, C., Poppa, T., Chen, K.H., Anderson, S.W., Damasio, H., Love, T.,  
1005 Hickok, G., 2015. Speech repetition as a window on the neurobiology of auditory-  
1006 motor integration for speech: A voxel-based lesion symptom mapping study.  
1007 *Neuropsychologia* 71, 18–27.
- 1008 Schneider, D.M., Nelson, A., Mooney, R., 2014. A synaptic and circuit basis for  
1009 corollary discharge in the auditory cortex. *Nature* 513, 189–194.

- 1010 Shuster, L.I., Lemieux, S.K., 2005. An fMRI investigation of covertly and overtly  
1011 produced mono- and multisyllabic words. *Brain and Language* 93, 20–31.
- 1012 Simmonds, A.J., Leech, R., Collins, C., Redjep, O., Wise, R.J.S., 2014a. Sensory-  
1013 Motor Integration during Speech Production Localizes to Both Left and Right Planum  
1014 Temporale. *Journal of Neuroscience* 34, 12963–12972.
- 1015 Simmonds, A.J., Wise, R.J.S., Collins, C., Redjep, O., Sharp, D.J., Iverson, P., Leech,  
1016 R., 2014b. Parallel systems in the control of speech. *Human Brain Mapping* 35, 1930–  
1017 1943.
- 1018 Simmonds, A.J., Wise, R.J.S., Dhanjal, N.S., Leech, R., 2011. A comparison of  
1019 sensory-motor activity during speech in first and second languages. *Journal of*  
1020 *Neurophysiology* 106, 470–478.
- 1021 Tachibana, R.O., Yanagida, M., Riquimaroux, H., 2010. Novel approach for  
1022 understanding the neural mechanisms of auditory-motor control: Pitch regulation by  
1023 finger force. *Neuroscience Letters* 482, 198–202.
- 1024 Tian, X., Poeppel, D., 2013. The Effect of Imagination on Stimulation: The  
1025 Functional Specificity of Efference Copies in Speech Processing. *Journal of Cognitive*  
1026 *Neuroscience* 25, 1020–1036.
- 1027 Tian, X., Poeppel, D., 2015. Dynamics of Self-monitoring and Error Detection in  
1028 Speech Production: Evidence from Mental Imagery and MEG. *Journal of Cognitive*  
1029 *Neuroscience* 27, 352–364.
- 1030 Timm, J., SanMiguel, I., Keil, J., Schröger, E., Schönwiesner, M., 2014. Motor  
1031 Intention Determines Sensory Attenuation of Brain Responses to Self-initiated  
1032 Sounds. *Journal of Cognitive Neuroscience* 26, 1481–1489.
- 1033 Tourville, J.A., Reilly, K.J., Guenther, F.H., 2008. Neural mechanisms underlying  
1034 auditory feedback control of speech. *Neuroimage* 39, 1429–1443.
- 1035 Tremblay, S., Shiller, D.M., Ostry, D.J., 2003. Somatosensory basis of speech  
1036 production. *Nature* 423, 866–869.
- 1037 Wikman, P.A., Vainio, L., Rinne, T., 2015. The effect of precision and power grips on  
1038 activations in human auditory cortex. *Frontiers in Neuroscience* 9, 378.
- 1039 Wilson, S.M., Iacoboni, M., 2006. Neural responses to non-native phonemes varying  
1040 in producibility: evidence for the sensorimotor nature of speech perception.  
1041 *Neuroimage* 33, 316–325.
- 1042 Woods, D.L., Alain, C., 2009. Functional imaging of human auditory cortex. *Current*  
1043 *Opinion in Otolaryngology and Head and Neck Surgery* 17, 407–411.
- 1044 Zhang, L., Xi, J., Xu, G., Shu, H., Wang, X., Li, P., 2011. Cortical dynamics of  
1045 acoustic and phonological processing in speech perception. *PLoS ONE* 6.

1046 Zvyagintsev, M., Clemens, B., Chechko, N., Mathiak, K.A., Sack, A.T., Mathiak, K.,  
1047 2013. Brain networks underlying mental imagery of auditory and visual information.  
1048 European Journal of Neuroscience 37, 1421–1434.